University of Southern Denmark
IMADA
Arthur Zimek
Jonatan Møller Gøttcke, Jonas Herskind Sejr

DM566/DM868/DM870: Data Mining and Machine Learning
Spring term 2019

Exercise 16: Group Project 2: Classification

Exercise 16-1 Evaluation of Classification Algorithms

You should work on this project in the same group as for project 1.

Start early with your exploration and plan sufficient time. Exercises 14–15 should give you additional insights and guidance along the way.

Choose data sets from the UCI machine learning repository `http://archive.ics.uci.edu/ml/` or some other classification problem that you find interesting and perform an empirical comparison of classification algorithms discussed in the lecture. (You can choose a data mining tool of your preference or your own implementation.)

Document your experiments. Guiding questions for your documentation could be:

- Which data sets did you choose and why?

- Which algorithms and which implementation did you use? (Why?)

- Report the achieved classification results (use several performance measures).

- How did you assess the performance?

- How did you determine parameters?

Discuss your experimental results. Guiding questions for the discussion could be:

- Which algorithm is most suitable for which dataset?

- Which algorithm is most unstable in terms of quality achieved when changing parameters?

- Is there an overall winner? (Regardless of whether the answer here is yes or no: how substantial and significant are your results w.r.t. this point?)

- Explain your findings!

Present your choices, experiments, results, and insights in a presentation in the exercise class in week 22 (prepare some slides for up to ten minutes presentation per group).